ENVIRONMENTAL HEALTH PERSPECTIVES

http://www.ehponline.org

Urinary Paraben Concentrations and Ovarian Aging among Women from a Fertility Center

Kristen W. Smith, Irene Souter, Irene Dimitriadis, Shelley Ehrlich, Paige L. Williams, Antonia M. Calafat, and Russ Hauser

http://dx.doi.org/10.1289/ehp.1205350

Received: 16 April 2012
Accepted: 1 August 2013
Advance Publication: 2 August 2013

National Institute of Environmental Health Sciences

Urinary Paraben Concentrations and Ovarian Aging among Women from a Fertility Center

Kristen W. Smith[1], Irene Souter[2], Irene Dimitriadis[1,2], Shelley Ehrlich[1], Paige L. Williams[3], Antonia M. Calafat[4], and Russ Hauser[1,2]

[1]Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, USA

[2]Department of Obstetrics and Gynecology, Division of Reproductive Endocrinology and Infertility, Harvard Medical School/Massachusetts General Hospital Fertility Center, Boston, Massachusetts, USA

[3]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA

[4]National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Address correspondence to:

Russ Hauser

665 Huntington Avenue

Building I, 14th Floor

Boston, MA 02115

Phone: 617-432-3326

rhauser@hsph.harvard.edu

Running Title: Parabens and Ovarian Aging

Acknowledgments and Grant Information: Work supported by grants ES009718, ES000002, and T32ES007069 from the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH). The authors gratefully acknowledge Xiaoyun Ye, Xiaoliu Zhou, Ryan Hennings, Amber Bishop, Tao Jia (CDC, Atlanta, GA) for measuring the urinary concentrations of the parabens. Disclaimer: The findings and conclusions in this report are those of the author and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Competing Financial Interests: None

Abstract

Background: Parabens are preservatives commonly used in personal care products, pharmaceuticals, and foods. There is documented widespread human exposure to parabens, and some experimental data suggesting they act as estrogenic endocrine disruptors. As far as we are aware, no epidemiologic studies have assessed female reproductive health effects in relation to paraben exposure.

Objective: Evaluate the association of urinary paraben concentrations with markers of ovarian reserve in a prospective cohort study of women seeking fertility treatment at Massachusetts General Hospital, Boston, MA.

Methods: Measures of ovarian reserve were day 3 follicle stimulating hormone (FSH), antral follicle count (AFC), and ovarian volume. In spot urine samples, collected prior to the assessment of outcome measures, we measured paraben concentrations (methyl-(MP), propyl-(PP), and butyl-paraben (BP)). We used linear and Poisson regression models to estimate associations of urinary paraben concentrations (in tertiles) with ovarian reserve measures.

Results: 192 women enrolled in 2004-2010 had at least one ovarian reserve outcome measured (mean age ±SD =36.1 ±4.5, range=21.0-46.7 years). MP and PP were detected in >99% of urine samples, and BP in >75%. There was a suggestive trend of lower AFC with increasing urinary PP tertiles (mean % change (95% CI) for tertiles 2 and 3, compared to tertile 1, respectively, were: -5.0 (-23.7, 18.4) and -16.3 (-30.8, 1.3), trend p-value=0.07) as well as higher day 3 FSH with higher urinary PP tertiles (mean change (95% CI) for tertiles 2 and 3, compared to tertile 1, were: 1.16 IU/L (-0.26, 2.57) and 1.02 IU/L (-0.40, 2.43), trend p-value=0.16). There was no consistent evidence of associations between urinary MP or BP and day 3 FSH or AFC, or between urinary MP, PP, or BP and ovarian volume.

Conclusions: Propyl-paraben may be associated with diminished ovarian reserve. However, our results require confirmation in further studies.

Introduction

Parabens are a family of chemicals commonly used as antimicrobial preservatives in personal care products, pharmaceuticals, and foods (Andersen 2008; NTP 2005; Orth 1980). Exposure to parabens can occur through ingestion, inhalation, or dermal absorption. Following excretion, the parent compounds can be measured in urine and have been shown to be valid biomarkers of exposure (Ye et al. 2006a).

Although parabens are quickly eliminated from the body (Janjua et al. 2008), parabens have been detected in the general US population (Calafat et al. 2010; Ye et al. 2006a). Methyl- (MP) and propyl-paraben (PP) are the two most commonly used parabens (Soni et al. 2005) and were detected in over 92% of a representative sample of the US population in the National Health and Nutrition Examination Survey (NHANES), whereas butyl-paraben (BP) was detected in 47% of NHANES participants (Calafat et al. 2010). Parabens have been detected in urine samples collected from infants (Calafat et al. 2009) and older children (Calafat et al. 2010; Casas et al. 2011; Wolff et al. 2010), in adults of reproductive age and older (Calafat et al. 2010; Meeker et al. 2011), and in pregnant women (Casas et al. 2011; Philippat et al. 2012; Smith et al. 2012), suggesting that exposure to parabens is ubiquitous and may begin in early life and extend throughout the lifespan.

Parabens are suspected endocrine disrupting chemicals that are estrogenic (Golden et al. 2005; Routledge et al. 1998; Soni et al. 2005), although they have a lower estrogen receptor binding affinity than endogenous estrogen (Routledge et al. 1998; Vo et al. 2010). Parabens have been shown to bind to both estrogen receptor (ER)-α and ER-β (Gomez et al. 2005; Okubo et al.

2001). The estrogenic activity of parabens increases with increasing length and branching of the alkyl chain (e.g. BP > PP > MP) (Byford et al. 2002; Routledge et al. 1998; Vo et al. 2010).

Based on available toxicologic data, MP and PP were classified as generally regarded as safe (GRAS) in 1972 by the US Food and Drug Administration (FDA) (US FDA 2006). A report from 2008 reviewed by the Cosmetic Ingredient Review panel concluded that parabens used in cosmetics, including BP, do not pose a safety risk based on the available data (Andersen 2008).

A few recent animal toxicity studies have reported adverse effects of some parabens on female reproductive and endocrine function (Kang et al. 2002; Taxvig et al. 2008; Vo et al. 2010). In one study evaluating pre-pubertal female rats treated orally with parabens, effects included, but were not limited to, a decrease in ovarian weight and histopathological changes in the ovaries, as well as altered estradiol and tetra-iodothyronine (T4), but not thyroid stimulating hormone (TSH) levels (Vo et al. 2010). Effects were seen with MP, BP, isopropyl-, and isobutyl-paraben, and the relationships varied by outcome, some of which were dose-dependent (Vo et al. 2010).

In a study evaluating pregnant rats exposed subcutaneously to parabens, a decrease in ER-β expression in the ovaries from female fetuses was observed among rats treated with BP (ER-β gene expression was significantly decreased for both BP doses administered compared to the control, but it is unclear if gene expression differed between the two BP doses) (Taxvig et al. 2008). However, no change in ovarian estradiol levels or ovarian histopathology was observed (Taxvig et al. 2008). The same study found no association of maternal or fetal reproductive hormone levels with ethyl-paraben or BP exposure. In pregnant rats exposed subcutaneously to BP, there was no evidence of effects on the reproductive organ weights and no histopathological abnormalities observed in the female offspring (Kang et al. 2002).

Overall, these limited studies suggest that some parabens may exert adverse endocrine disrupting effects on female animals, but additional toxicologic data, including mechanistic studies are needed. Human data on the reproductive health effects of paraben exposure are limited, and, as far as we are aware, no studies have reported on the association of urinary paraben concentrations with female reproductive health outcomes.

Given the suspected endocrine disrupting properties of parabens and the sensitivity of oogenesis to proper estrogen signaling, we were interested in evaluating the potential association between urinary paraben concentrations and markers of ovarian reserve. Hormonal and ultrasonographic markers of ovarian reserve are commonly used by reproductive endocrinology and infertility specialists to evaluate a woman's response to ovarian stimulation and include follicle stimulating hormone (FSH) serum concentration on day 3 of the menstrual cycle, antral follicle count (AFC), and ovarian volume (OV). Typically, as a woman's age increases, her ovarian reserve diminishes, which is referred to as "ovarian aging" and is associated with reduced fertility. Among women undergoing Assisted Reproductive Technology (ART), ovarian aging is also associated with a decreased response to ovarian stimulation protocols and lower pregnancy success rates (Elter et al. 2005; Levi et al. 2001). This diminished ovarian reserve is generally indicated by higher day 3 FSH levels and a lower AFC and OV. However, there are other factors besides age that could be associated with a diminished ovarian reserve, including, possibly, exposure to endocrine disrupting chemicals. Therefore, the objective of this study was to evaluate if exposure to parabens, assessed from urinary paraben concentrations, is associated with diminished ovarian reserve among women undergoing *in vitro* fertilization (IVF) or intrauterine insemination (IUI).

Methods

Participants. Study participants included female patients from the Massachusetts General Hospital (MGH) Fertility Center undergoing infertility evaluation and participants in our ongoing prospective cohort study on environmental risk factors for reproductive health (Environment and Reproductive Health Study). The participants had at least one hormonal or ultrasonographic marker of ovarian reserve measured (day 3 FSH, AFC, or OV) and also contributed at least one urine sample for the measurement of paraben concentrations prior to the measurement of the markers of ovarian reserve. All female patients over 18 years of age and less than 46 years (at enrollment) seeking infertility evaluation or treatment at the MGH Fertility Center were eligible to participate (close to 100% of patients are eligible) and approximately 60% consented. We excluded participants that had an oophorectomy (N=5). We recruited participants between December 2004 and October 2010 and followed them from study entry until the discontinuation of fertility treatment, a live birth, or loss to follow-up. Two patients re-enrolled in the study after the end of the initial follow-up period and only data from their first enrollment were included in this analysis. The study was approved by the Human Studies Institutional Review Boards of the MGH, Harvard School of Public Health (HSPH), and the Centers for Disease Control and Prevention (CDC). Participants signed an informed consent after the study procedures were explained by a research nurse and all questions were answered.

Clinical data. Clinical information was abstracted from the patient's electronic medical record by a research nurse. An intravenous blood sample was drawn on the third day of the menstrual cycle and the serum was analyzed for FSH with an automated electrochemiluminescence immunoassay at the MGH Core Laboratory as previously described (Mok-Lin et al. 2010). AFC and OV were measured for both ovaries by a reproductive endocrinology and infertility specialist at the MGH Fertility Center using transvaginal ultrasound. We calculated the OV using the following formula: (length (mm) x width (mm) x height (mm)) x (π / 6). We used the sum of antral follicles from the left and right ovaries (antral follicle count, AFC) and the average ovarian volume (of left and right ovaries, OV) in the analysis. Subsequent to an infertility evaluation, each patient was given an infertility diagnosis by a physician at the MGH Fertility Center according to the Society for Assisted Reproductive Technology (SART) definitions as previously described (Mok-Lin et al. 2010). The participant's date of birth, weight, and height were collected using a nurse administered questionnaire at entry into the study (weight and height were measured by the nurse).

Urinary paraben measurements. We collected a convenience spot urine sample from the women at the time of recruitment and at subsequent visits during infertility treatment cycles. Although participants were recruited into this study beginning in 2004, the measurement of parabens in urine did not begin until August 2005, when these chemicals were added to the study protocol. We collected samples between August 2005 and November 2010. Urine was collected in a sterile polypropylene cup. After measuring specific gravity (SG) using a handheld refractometer (National Instrument Company, Inc., Baltimore, MD, USA), the urine was divided into aliquots and frozen at -80°C. Samples were shipped on dry ice overnight to the CDC (Atlanta, GA, USA), where concentrations of total (free + conjugated) MP, PP, and BP were measured using on-line solid phase extraction-high performance liquid chromatography-isotope dilution tandem mass spectrometry as previously reported (Ye et al. 2006b). Standard QA/QC procedures were followed (US CDC 2010). The limits of detection (LOD) were 1.0 µg/L for MP and 0.2 µg/L for PP and BP.

Statistical analysis. Demographic characteristics of the study participants (mean and percentage) were reported separately for each outcome measure because the number of participants varied by outcome, as well as for participants with any ovarian reserve outcome measured. The distribution of day 3 FSH, AFC, and OV were described using the mean and standard deviation (SD), median and interquartile range (IQR), and range. We computed the within-person geometric mean (GM) of all urinary paraben concentrations (MP, PP, and BP) measured prior to outcomes as a summary exposure measure for each participant. We summarized the distribution of exposures using the median, interquartile range, and range of paraben urinary concentrations. We assigned urinary concentrations below the LOD with a value equal to the LOD divided by the square root of two (Hornung and Reed 1990). We corrected urinary paraben concentrations for SG using a modification of a previously described formula (Duty et al. 2005): $Pc = P[(1.016 - 1)/SG - 1]$, where Pc is the SG-corrected paraben concentration (µg/L), P is the measured paraben concentration (µg/L), and 1.016 is the mean (and median) SG level in the study population. We used SG-corrected paraben concentrations in all analyses. We calculated the Spearman correlation between the within-person GMs of the different parabens.

We calculated the Spearman correlation between the markers of ovarian reserve (day 3 FSH, AFC, and OV), age, and body mass index (BMI). Among women with both outcome measures we calculated the correlation between FSH and AFC, FSH and OV, and AFC and OV. We calculated the correlation of age and BMI with each separate outcome measure using all available measurements. We were interested in evaluating the association between BMI and these measures of ovarian reserve because a higher BMI has been shown to be associated with infertility (Pasquali et al. 2003; Shah et al. 2011), although there is limited evidence in the literature for an association between ovarian reserve measures and BMI (Su et al. 2008).

We used multivariable linear regression to estimate associations between within-person MP, PP, and BP GM concentrations (divided into tertiles) with day 3 FSH and OV. OV was natural log-transformed prior to all regression analyses to reduce skewness. We used Poisson regression to estimate associations between within-person MP, PP, and BP GM concentrations (divided into tertiles) with AFC. Covariates considered for inclusion in the regression models included age in years at the time of the outcome measure and BMI (kg/m^2) at entry into the study (both modeled as continuous measures), and were included when related to the outcome measure in univariate regression models ($p<0.20$). We also considered age categorized into <37 years and ≥37 years for inclusion as a covariate in a sensitivity analysis because the ability to become pregnant declines around age 37 among women in the US population undergoing ART (US CDC, 2010). To allow for easier interpretation of the results, we exponentiated the parameter estimates for the linear regression model evaluating log-transformed OV and for the Poisson regression model evaluating log AFC. The mean percent change in the outcome from the lowest tertile of paraben concentrations is presented for these two outcomes (OV and AFC). We conducted all tests for trend by assigning each urinary paraben concentration tertile an ordinal integer value of 0 (lowest tertile) to 2 (highest tertile).

As a sensitivity analysis, we reran the regression models for AFC and OV excluding patients diagnosed with polycystic ovarian syndrome (PCOS) because these women tend to have a higher AFC and larger OV compared to women without this disease. PCOS was defined using a SART diagnosis (primary, secondary, or tertiary) of ovulatory disorder.

In a secondary analysis, we combined the parabens using two methods. We first used an estrogen equivalency (EEQ) factor approach (Safe 1998; Shirai et al. 2012) using the following formula: EEQ(parabens) = $MP_m*1 + PP_m*83.3 + BP_m*250$, where MP_m, PP_m, and BP_m are the SG-adjusted within-person GM molar concentrations and the potency factors come from an *in vitro* yeast based estrogen assay (BP ~10,000 times less potent than 17β estradiol, PP ~30,000 times less potent, and MP ~2,500,000 less potent) (Routledge et al. 1998). Second, we summed the urinary paraben concentrations using the following formula: Sum(parabens) = $MP_m + PP_m + BP_m$. We used multivariable linear regression to separately evaluate the association between EEQ(parabens) and Sum(parabens) (both divided into tertiles) with day 3 FSH and OV. OV was natural log-transformed prior to all regression analyses to reduce skewness. We used Poisson regression to analyze the association between EEQ(parabens) and Sum(parabens) (both divided into tertiles) with AFC. We conducted all statistical analyses using SAS version 9.2 (SAS Institute Inc., Cary, NC) and considered two-sided significance levels less than 0.05 as statistically significant.

This exploratory study provided 80% power to detect a difference of 2.11 IU/L in day 3 FSH or 0.38mm^3 in log(OV), for comparing high or medium paraben urinary concentrations versus low paraben urinary concentrations (0.67 SDs). Similarly, the study design provided 80% power for detecting a difference in AFC of 4.2, corresponding to a decrease of ~38% between high versus low paraben urinary concentrations.

Results

A total of 193 women had at least one measure of ovarian reserve available and at least one urinary paraben concentration measurement. We excluded one woman with all urine samples missing SG measurements, resulting in a final sample of 192 women. Because all outcome measures were not available for all participants, we evaluated each outcome separately using all available measurements: day 3 FSH was measured in 110 women, AFC in 142 women, and OV in 109 women. There were 44 women with all three measures, 81 women with two measures, and 67 women with only one measure available. We collected between 1 and 14 urine samples from each participant (median in each dataset=1 sample/participant; range of means in all three datasets=2.2 - 2.6 samples/participant) that contributed to the GM summary exposure measure. The urine samples were collected between 0 (the same day) and 1,145 days prior to the outcome measure for the AFC dataset, and between 0 and 981 days prior for the FSH and OV datasets. The mean (±SD) and median days collected prior to the outcome measure are as follows: mean=142 ± 182, median=94 (AFC); mean=157±159), median=108 (FSH); mean=110±128), median=77 (OV).

Women were primarily Caucasian, non-smokers, over 35 years of age, and had a mean BMI of 25.4 ± 5.15 (Table 1). The SART diagnosis was most commonly female factor, followed by male factor and unexplained infertility (Table 1). There was no significant difference in age or the number of participants diagnosed with PCOS in each of the three outcome subgroups (data not shown). The mean day 3 FSH (IU/L) was 7.39 ± 3.17 with a range of 0.10 to 26.0. The median (IQR) AFC sum (left and right ovaries) was 11 (7 - 15) with a range of 2 to 40. The median (IQR) OV (mm^3) was 4,928 (3,634 - 7,588) with a range of 1,359 to 27,834.

Urinary paraben concentrations were similar to those in the general population; MP and PP were detected in greater than 99% of samples, and BP in greater than 75% (Table 2). In each of the three datasets, there was a strong correlation between concentrations of MP and PP (Spearman r range = 0.81-0.85) and a moderate correlation for MP and BP (Spearman r range =0.40-0.47) and PP and BP (Spearman r range = 0.43-0.46).

The correlation of day 3 FSH, AFC, and OV was assessed among women having both outcome measures for each pair of correlations. Among women with both FSH and AFC (N=85), FSH was negatively correlated with AFC (r=-0.40, p=0.002). Among women with both FSH and OV (N=49), FSH was negatively correlated with OV (r=-0.36, p=0.01). Among women with both AFC and OV (N=79), AFC was positively correlated with OV (r=0.47, p<0.0001). FSH was positively correlated with age (r=0.29, p=0.002), but not correlated with BMI (r=-0.048, p=0.62). AFC and OV were negatively correlated with age (r=-0.44, p < 0.001; r=-0.21, p=0.025, respectively), but not correlated with BMI (r=0.036, p=0.67; r=0.040, p=0.68, respectively).

PP concentration was positively related to day 3 FSH, with mean day 3 FSH higher in both tertiles 2 and 3, compared to tertile 1, although there was not a significantly increasing trend across tertiles, with the mean level in tertile 3 similar to tertile 2 (trend test p-value=0.16). For MP and BP, mean day 3 FSH was also higher in both tertiles 2 and 3, compared to tertile 1, although the mean level was lower in tertile 3 compared to tertile 2 for both analytes (trend test p-values for MP and BP, respectively=0.64 and 0.60) (Table 3). There was a suggestive trend for lower AFC among women with higher PP concentrations, with the mean % difference from tertile 1 in AFC decreasing across tertiles (trend test p-value=0.07). For MP, the magnitude of the parameter estimates in tertiles 2 and 3 was similar to PP, although the trend was not significant (trend test p-value=0.31). No association was observed with BP and AFC (Table 4). There was no evidence of an association between urinary MP, PP, or BP concentrations and OV (Table 5). Age was significantly negatively associated with AFC and OV and positively associated with FSH and was included as a covariate in all regression models as a continuous measure (age was not associated with the exposures), whereas BMI was not observed to be associated (p > 0.20) and was not included.

In a sensitivity analysis controlling for age categorized into <37 years and $\geq$ 37 years, the association between day 3 FSH and PP became stronger with a mean difference (95% CI) in day 3 FSH of 1.24 IU/L (-0.18, 2.67 IU/L) in the 3rd tertile (using age as a continuous covariate, mean difference (95% CI)= 1.02 (-0.40, 2.43)), and of 1.19 IU/L (-0.22, 2.60 IU/L) in the 2nd tertile (using age as a continuous covariate, mean difference (95% CI)= 1.16 (-0.26, 2.57)), both compared to the lowest tertile (with the trend test p-value becoming borderline significant, p-value = 0.08). Including the categorized age variable, the association between day 3 FSH and MP became stronger with a mean difference (95% CI) in day 3 FSH of 0.57 IU/L (-0.85, 1.99 IU/L) in the 3rd tertile, and of 1.21 IU/L (-0.21, 2.63 IU/L) in the 2nd tertile, both compared to the lowest tertile. Including the categorized age variable, the association between day 3 FSH and BP became stronger with a mean difference (95% CI) in day 3 FSH of 0.53 IU/L (-0.89, 1.95 IU/L) in the 3rd tertile, and of 1.10 IU/L (-0.32, 2.52 IU/L) in the 2nd tertile, both compared to the lowest tertile. The trend test p-values for MP and BP remained non-significant when including the categorized age variable. The relationship of MP, PP, and BP with AFC and OV was similar when including the categorized age variable as compared to including the continuous age variable (data not shown).

Among patients with an available SART diagnosis (N=1 missing from each dataset), when excluding PCOS patients (N=25 in AFC dataset; N=20 in OV dataset) and controlling for age (continuous), the relationship of PP with AFC was attenuated (data not shown). Excluding PCOS patients altered parameter estimates for associations of MP and BP with AFC and associations of MP, PP, and BP with OV, but estimates were imprecise and remained non-significant (data not shown).

In a secondary exploratory analysis evaluating the association of combined concentrations of parabens with the three outcomes (controlling for age as a continuous measure), we found that EEQ(parabens) was negatively related to AFC in the 3rd tertile, with a mean difference of -6.4% (-24.4%, 15.8%), and little evidence of an association in the 2nd tertile, with a mean difference of 1.1% (-17.3%, 23.5%), both compared to the lowest tertile, although the trend was not statistically significant (trend test p-value = 0.54) (Supplemental Material, Table S1). Similarly, we found that Sum(parabens) was negatively related to AFC, with a mean difference of -10.8% (-28.2%, 10.7%) in the 3rd tertile, and of -6.5% (-23.4%, 14.1%) in the 2nd tertile, both compared to the lowest tertile, although the trend was not statistically significant (trend test p-value = 0.30). We found no significant relationships between EEQ(parabens) or Sum(parabens) with day 3 FSH or OV (see Supplemental Material, Tables S2 and S3, respectively).

Discussion

As far as we are aware, this is the first epidemiologic study to assess female reproductive health outcomes in relation to biomarkers of paraben exposure. We found that close to 100% of the women included in this study had detectable urinary concentrations of MP and PP, while over 75% of women had detectable concentrations of BP. Paraben urinary concentrations were similar to those reported for all females from NHANES in 2005-2006 (Calafat et al. 2010). Median (IQR) unadjusted paraben urinary concentrations for women from NHANES in 2005-2006 were as follows: 137 µg/L (35.4, 356 µg/L), 29.1 µg/L (5.30, 93.0 µg/L), and 0.50 µg/L (<LOD, 3.70 µg/L) for MP, PP, and BP, respectively (Calafat et al. 2010). We found suggestive evidence of a negative relationship of urinary PP with AFC, considered one of the best markers of ovarian reserve (Rosen et al. 2012). Higher urinary PP was associated with a higher day 3 FSH which is

consistent with PP's negative association with AFC. The positive relationship of urinary PP with day 3 FSH approached statistical significance when controlling for age categorized into <37 years and ≥37 years, although the magnitude of the association was similar as when controlling for age as a continuous measure. These findings suggest that exposure to PP may adversely affect ovarian reserve, and thus contribute to ovarian aging, among women attending a fertility clinic. Although there was suggestive evidence of a negative relationship between MP and AFC, there was no clear evidence of associations between urinary MP or BP concentrations with any of the markers of ovarian reserve. Similar to the relationship between urinary PP with AFC, in an exploratory analysis the three parabens combined were negatively related to AFC using both EEQ(parabens) and Sum(parabens), although the relationship did not approach statistical significance.

It has been established that parabens are estrogenic (Golden et al. 2005; Routledge et al. 1998; Soni et al. 2005), and that they bind to both ER-α and ER-β (Gomez et al. 2005; Okubo et al. 2001). Although the estrogen receptor binding affinity of parabens is much lower than that of endogenous estrogen (Darbre and Harvey 2008), oogenesis is highly dependent upon proper estrogen signaling (Hewitt et al. 2005), and therefore even slight changes in the ovarian hormonal environment (either *in utero* or later in life) could contribute to altered ovarian function. The relationship of PP with diminished ovarian reserve is consistent with the animal data showing that the estrogenicity of parabens, and therefore the potential for reproductive toxicity, is greater in PP compared to MP (Byford et al. 2002; Routledge et al. 1998; Vo et al. 2010). Although the animal data also show that BP is more estrogenic than PP or MP, BP was not detected as frequently and, when detected, BP urinary concentrations were much lower than either PP or MP in this study, which may explain the lack of an association of BP with markers of ovarian reserve. It is also possible that biological activity and mechanisms of action differ between the parabens, although, as far as we know, this has not been studied.

There are a few studies, conducted in female rats and mice, suggesting an association between paraben exposure and reproductive outcomes. These include changes in ovarian weight and histopathology, as well as changes in ER-α and ER-β gene expression. In a study on pre-pubertal female mice treated orally with parabens, adverse effects included a decrease in the ovarian weight (from MP and isopropyl-paraben, but not PP or BP) and histopathological changes in the ovaries (from MP, isopropyl-paraben, BP, and isobutyl-paraben, but not PP) (Vo et al. 2010). The histopathological changes noted included a decrease of the corpora lutea, an increase in the number of cystic follicles, and a thinning of the follicular cells, which suggests that post-natal paraben exposure could adversely influence ovarian follicle development and thus potentially lead to diminished ovarian reserve. These changes could be due to the estrogenic action of parabens. Similar effects have been observed in adult mice exposed to diethylstilbestrol (Hong et al. 2010).

Another study, looking at pregnant rats exposed subcutaneously to parabens, found a decrease in ER-β expression in the ovaries from female fetuses (from BP, although MP and PP were not evaluated) (Taxvig et al. 2008). An *in vitro* study using human breast cancer MCF-7 cells also found that ER-α expression decreased and progesterone receptor (PR) expression increased after administration of BP and isobutyl-paraben (MP and PP were not evaluated) (Okubo et al. 2001). It is possible that altered gene expression related to *in utero* paraben exposure could adversely affect the ovarian follicle pool. It is well known that proper estrogen signaling is a key component in the development of the ovarian follicle pool *in utero* (Crain et al. 2008), and when disrupted could manifest as diminished ovarian reserve during a woman's reproductive years.

A previous study including the same patients from the MGH Fertility Center found that one urine sample was reasonably representative of urinary paraben concentrations over several months (intraclass correlation coefficients between 0.4-0.5 for MP, PP, and BP using non-pregnancy samples) (Smith et al. 2012). In the current study, because multiple samples were collected from some women, a summary exposure measure was used for each participant by taking the geometric mean of all urine samples collected prior to the outcome measure. Although one urine sample may reasonably represent several months exposure, one strength of the present study is that the collection of multiple samples should reduce exposure misclassification during that time period. However, a limitation of this study is that the time period of collection of the urine samples was up to three years prior to the outcome measure. It is unknown if the window of exposure that is most etiologically relevant to the outcomes assessed is the year prior to the outcome measure, for example, or any earlier period in the life course (e.g. pubertal or even *in utero* exposure). If paraben exposure within several months prior is the relevant window of exposure, the summary exposure measure used may reasonably represent the relevant exposure period. However, any exposure misclassification in this study is thought to be non-differential. In future studies with larger sample sizes, we recommend exploration of whether samples collected closer in time (i.e., 3 month window) are more strongly associated with the outcome measures compared to samples collected more remotely.

Another limitation of this study is the relatively small sample size, which may limit our ability to detect an association. Also, not all women had all three of the outcome measures as they are all not always clinically performed. However, this study is the first of its kind and further investigation using a larger sample size is suggested to detect potentially subtle changes in markers of ovarian reserve in response to these suspected endocrine disrupting chemicals. The high proportion of Caucasians and older women, as well as the sole inclusion of women from a fertility clinic undergoing IVF or IUI, with varied SART diagnoses, in our study may also limit the generalizability of these findings to non-Caucasians, younger women, and women with no difficulties conceiving. In addition, because there are numerous xenoestrogens in personal care products, food, and medications (e.g. parabens, bisphenol A, benzophenone-3, triclosan) it is suggested that future studies take into account the potential for the effect of estrogenic mixtures (e.g. assessing interactions between exposure categories of the chemicals). Finally, there is also the possibility of bias from uncontrolled confounding given that personal care product use may change with age.

In conclusion, this study provides suggestive evidence that exposure to PP may lead to diminished ovarian reserve and contribute to ovarian aging among women at an infertility clinic. It has been estimated that in 2002 there were over 7 million women with impaired fecundity in the US, and over 5 million women were reported as seeking help to become pregnant (Chandra et al. 2005). This is a large sub-population of women that may be especially sensitive to endocrine disrupting chemicals. Finally, although the parabens evaluated for this study are considered to be safe (have a GRAS designation) based on a 1972 decision by the US FDA, given their widespread use and ubiquitous human exposure, further research using modern toxicological designs and endpoints may be warranted. In addition, our results suggest the need for future human studies to explore these associations in other populations with a larger sample size.

References

Andersen FA. 2008. Final amended report on the safety assessment of methylparaben, ethylparaben, propylparaben, isopropylparaben, butylparaben, isobutylparaben, and benzylparaben as used in cosmetic products. Int J Toxicol 27 Suppl 4:1-82.

Byford JR, Shaw LE, Drew MG, Pope GS, Sauer MJ, Darbre PD. 2002. Oestrogenic activity of parabens in MCF7 human breast cancer cells. J Steroid Biochem Mol Biol 80(1):49-60.

Calafat AM, Ye X, Wong LY, Bishop AM, Needham LL. 2010. Urinary concentrations of four parabens in the U.S. population: NHANES 2005-2006. Environ Health Perspect 118(5):679-685.

Calafat AM, Weuve J, Ye X, Jia LT, Hu H, Ringer S, et al. 2009. Exposure to bisphenol A and other phenols in neonatal intensive care unit premature infants. Environ Health Perspect 117(4):639-644.

Casas L, Fernandez MF, Llop S, Guxens M, Ballester F, Olea N, et al. 2011. Urinary concentrations of phthalates and phenols in a population of Spanish pregnant women and children. Environ Int 37(5):858-866.

CDC (Centers for Disease Control and Prevention). 2010. 2008 Assisted Reproductive Technology Success Rates, National Summary and Fertility Clinic Reports. Available: http://www.cdc.gov/art/ART2008/PDF/ART_2008_Full.pdf [accessed 2 February 2012].

CDC (Centers for Disease Control and Prevention). 2010. Laboratory Procedure Manual. Available: http://www.cdc.gov/nchs/data/nhanes/nhanes_07_08/EPH_E_met_phenols_parabens.pdf [accessed 18 July 2012].

Chandra A, Martinez G, Mosher W, Abma J, Jones J. 2005. Fertility, family planning, and reproductive health of U.S. women: Data from the 2002 National Survey of Family Growth. National Center for Health Statistics. Vital Health Stat 23(25). Available: http://www.cdc.gov/nchs/data/series/sr_23/sr23_025.pdf [accessed 18 May 2011].

Crain DA, Janssen SJ, Edwards TM, Heindel J, Ho SM, Hunt P, et al. 2008. Female reproductive disorders: the roles of endocrine-disrupting compounds and developmental timing. Fertil Steril 90(4):911-940.

Darbre PD, Harvey PW. 2008. Paraben esters: review of recent studies of endocrine toxicity, absorption, esterase and human exposure, and discussion of potential human health risks. J Appl Toxicol 28(5):561-578.

Duty SM, Ackerman RM, Calafat AM, Hauser R. 2005. Personal care product use predicts urinary concentrations of some phthalate monoesters. Environ Health Perspect 113(11):1530-1535.

Elter K, Kavak ZN, Gokaslan H, Pekin T. 2005. Antral follicle assessment after down-regulation may be a useful tool for predicting pregnancy loss in in vitro fertilization pregnancies. Gynecol Endocrinol 21:33-37.

FDA (Food and Drug Administration). 2006. GRAS Substances (SCOGS) Database. Available: http://www.fda.gov/Food/FoodIngredientsPackaging/GenerallyRecognizedasSafeGRAS/GRASSubstancesSCOGSDatabase/default.htm [accessed 13 December 2011].

Golden R, Gandy J, Vollmer G. 2005. A review of the endocrine activity of parabens and implications for potential risks to human health. Crit Rev Toxicol 35(5):435-458.

Gomez E, Pillon A, Fenet H, Rosain D, Duchesne MJ, Nicolas JC, et al. 2005. Estrogenic activity of cosmetic components in reporter cell lines: parabens, UV screens, and musks. J Toxicol Environ Health A 68(4):239-251.

Hewitt SC, Harrell JC, Korach KS. 2005. Lessons in estrogen biology from knockout and transgenic animals. Annu Rev Physiol 67:285-308.

Hong Y, Wang J, Zhang P, Yang S, Song K, Yu F, et al. 2010. Histopathological and gene expression analysis of mice exposed to diethylstilbestrol. Toxicol Mech Methods 20(3):105-11.

Hornung R, Reed L. 1990. Estimation of average concentration in the presence of nondetectable values. App Occup Environ Hyg 5:46-51.

Janjua NR, Frederiksen H, Skakkebaek NE, Wulf HC, Andersson AM. 2008. Urinary excretion of phthalates and paraben after repeated whole-body topical application in humans. Int J Androl 31(2):118-130.

Kang KS, Che JH, Ryu DY, Kim TW, Li GX, Lee YS. 2002. Decreased sperm number and motile activity on the F1 offspring maternally exposed to butyl p-hydroxybenzoic acid (butyl paraben). J Vet Med Sci 64(3):227-235.

Levi AJ, Raynault MF, Bergh PA, Drews MR, Miller BT, Scott RT. 2001. Reproductive outcome in patients with diminished ovarian reserve. Fertil Steril 76:666–669.

Meeker JD, Yang T, Ye X, Calafat AM, Hauser R. 2011. Urinary concentrations of parabens and serum hormone levels, semen quality parameters, and sperm DNA damage. Environ Health Perspect 119(2):252-257.

Mok-Lin E, Ehrlich S, Williams PL, Petrozza J, Wright DL, Calafat AM, et al. 2010. Urinary bisphenol A concentrations and ovarian response among women undergoing IVF. Int J Androl 33(2):385-393.

National Toxicology Program (NTP). 2005. Butylparaben [CAS No. 94-26-8] Review of Toxicological Literature. Available: http://ntp.niehs.nih.gov/ntp/htdocs/chem_background/exsumpdf/butylparaben.pdf [accessed 3 January 2011].

Orth DS. 1980. Use of parabens as cosmetic preservatives. Int J Dermatol 19(9):504-505.

Okubo T, Yokoyama Y, Kano K, Kano I. 2001. ER-dependent estrogenic activity of parabens assessed by proliferation of human breast cancer MCF-7 cells and expression of ERalpha and PR. Food Chem Toxicol 39(12):1225-1232.

Pasquali R, Pelusi C, Genghini S, Cacciari M, Gambineri A. 2003. Obesity and reproductive disorders in women. Hum Reprod Update 9(4):359-372.

Philippat C, Mortamais M, Chevrier C, Petit C, Calafat AM, Ye X, et al. 2012. Exposure to phthalates and phenols during pregnancy and offspring size at birth. Environ Health Perspect 120(3):464-470.

Rosen MP, Johnstone E, McCulloch CE, Schuh-Huerta SM, Sternfeld B, Reijo-Pera RA, et al. 2012. A characterization of the relationship of ovarian reserve markers with age. Fertil Steril 97(1):238-243.

Routledge EJ, Parker J, Odum J, Ashby J, Sumpter JP. 1998. Some alkyl hydroxy benzoate preservatives (parabens) are estrogenic. Toxicol Appl Pharmacol 153(1):12-19.

Safe SH. 1998. Hazard and risk assessment of chemical mixtures using the toxic equivalency factor approach. Environ Health Perspect 106(Suppl 4):1051-1058.

Shah DK, Missmer SA, Berry KF, Racowsky C, Ginsburg ES. 2011. Effect of obesity on oocyte and embryo quality in women undergoing in vitro fertilization. Obstet Gynecol 118(1):63-70.

Shirai S, Suzuki Y, Yoshinaga J, Shiraichi H, Mizumoto Y. 2012. Urinary excretion of parabens in pregnancy Japanese Women. Repro Tox 35: 96-101.Smith KW, Braun JM, Williams PL, Ehrlich S, Correia KF, Calafat AM, et al. 2012. Predictors and Variability of Urinary Paraben Concentrations in Men and Women, Including before and during Pregnancy. Environ Health Perspect 120: 1538-1543.

Soni MG, Carabin IG, Burdock GA. 2005. Safety assessment of esters of p-hydroxybenzoic acid (parabens). Food Chem Toxicol 43(7):985-1015.

Su HI, Sammel MD, Freeman EW, Lin H, DeBlasis T, Gracia CR. 2008. Body size affects measures of ovarian reserve in late reproductive age women. Menopause 15(5):857-861.

Taxvig C, Vinggaard AM, Hass U, Axelstad M, Boberg J, Hansen PR, et al. 2008. Do parabens have the ability to interfere with steroidogenesis? Toxicol Sci 106(1):206-213.

Vo TT, Yoo YM, Choi KC, Jeung EB. 2010. Potential estrogenic effect(s) of parabens at the prepubertal stage of a postnatal female rat model. Reprod Toxicol 29(3):306-316.

Wolff MS, Teitelbaum SL, Pinney SM, Windham G, Liao L, Biro F, et al. 2010. Investigation of relationships between urinary biomarkers of phytoestrogens, phthalates, and phenols and pubertal stages in girls. Environ Health Perspect 118(7):1039-1046.

Ye X, Bishop AM, Reidy JA, Needham LL, Calafat AM. 2006a. Parabens as urinary biomarkers of exposure in humans. Environ Health Perspect 114(12):1843-1846.

Ye X, Kuklenyik Z, Bishop AM, Needham LL, Calafat AM. 2006b. Quantification of the urinary concentrations of parabens in humans by on-line solid phase extraction-high performance liquid chromatography-isotope dilution tandem mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci 844(1):53-59.

Table 1. Characteristics by ovarian reserve outcome of 192 women participants of a prospective fertility study at Massachusetts General Hospital enrolled between 2004-2010[a]

Characteristic	Day 3 FSH (Subject N=110)	Antral Follicle Count (Subject N=142)	Ovarian Volume (Subject N=109)	Any Ovarian Reserve Outcome Measured (N=192)
Age (years)				
Mean ± SD	36.1 ± 4.67	36.3 ± 4.24	35.6± 4.64	36.1±4.48[b]
Range	(22.0 - 45.3)	(21.0 - 44.8)	(21.7 - 46.7)	(21.0 - 46.7)
BMI (kg/m^2)[c]				
Mean ± SD	25.6 ± 5.54	25.3 ± 5.14	24.9 ± 4.70	25.4 ± 5.15
Range	(17.3 - 42.4)	(17.5 - 40.5)	(17.5 - 40.5)	(17.3 - 42.4)
Race [n(%)]				
Caucasian	85 (77)	115 (81)	92 (84)	156 (81)
African-American/Black	8 (7)	6 (4)	5 (5)	10 (5)
Asian	5 (5)	9 (6)	7 (6)	10 (5)
Native American/Alaska Native	1 (1)	1 (1)	0 (0)	1 (1)
Other	11 (10)	11 (8)	5 (5)	15 (8)
Smoking history [n(%)]				
Never	86 (78)	106 (75)	76 (70)	142 (74)
Former	20 (18)	29 (20)	27 (25)	42 (22)
Current	4 (4)	7 (5)	6 (5)	8 (4)
SART diagnosis [n(%)][d]				
Female factor	48 (44)	60 (43)	40 (37)	81 (42)
Endometriosis	8 (7)	8 (6)	8 (7)	12 (6)
Tubal factor	8 (7)	10 (7)	5 (5)	13 (7)
Diminished ovarian reserve	14 (13)	18 (13)	9 (8)	23 (12)
Ovulation disorders	17 (16)	23 (16)	18 (17)	31 (16)
Uterine disorders	1 (1)	1 (1)	0 (0)	2 (1)
Male factor	31 (28)	32 (23)	30 (28)	49 (26)
Unexplained	23 (21)	40 (28)	31 (29)	50 (26)
Other	7 (6)	9 (6)	7 (6)	11 (6)

[a]n=192 total unique participants

[b]For women with more than one outcome measure, the average age of the woman from each outcome measure was used to calculate the mean and standard deviation

[c]FSH dataset: n=109; OV dataset: n=108; Among all women with any outcome measured: n=191

[d]Primary SART diagnosis is reported; FSH dataset: n=109; AFC dataset: n=141; OV dataset: n=108; Among all women with any outcome measured: n=191

Table 2. Distribution of urinary paraben concentrations (μg/L) measured among participants of a prospective fertility study at Massachusetts General Hospital enrolled between 2004-2010, by ovarian reserve outcome

			Uncorrected			SG-corrected		
Paraben	**N**	**% Detect[a]**	**Min**	**Median (IQR)**	**Max**	**Min**	**Median (IQR)**	**Max**
Day 3 FSH								
MP	110	100%	6.70	210 (75.2, 520)	4400	10.8	249 (89.0, 549)	2428
PP	110	99.7%	0.20	49.6 (13.0, 89.3)	1000	0.46	55.1 (22.5, 124)	727
BP	110	80.0%	<LOD	2.08 (0.40, 6.58)	142	<LOD	2.83 (0.40, 9.60)	177
AFC								
MP	142	100%	3.00	180 (74.7, 400)	4400	5.13	227 (84.4, 492)	2428
PP	142	99.4%	<LOD	37.1 (17.4, 83.4)	1430	<LOD	52.1 (21.5, 110)	727
BP	142	78.4%	<LOD	1.42 (0.30, 6.07)	142	<LOD	1.60 (0.33, 9.49)	177
OV								
MP	109	100%	3.00	158 (57.7, 343)	4400	7.77	219 (72.2, 429)	2428
PP	109	99.2%	<LOD	35.5 (11.6, 88.7)	1430	<LOD	61.0 (14.6, 110)	654
BP	109	75.8%	<LOD	1.53 (0.30, 6.13)	142	<LOD	2.18 (0.45, 8.50)	102

Abbreviation: IQR = Interquartile range; LOD = Limit of detection

[a] % of concentrations above the LOD of MP=1μg/L, PP=0.2μg/L, and BP=0.2μg/L; Total samples with PP concentrations <LOD by dataset: 1 of 290 (FSH); 2 of 310 (AFC); 2 of 252 (OV); Total samples with BP concentrations <LOD by dataset: 58 of 290 (FSH); 67 of 310 (AFC); 61 of 252 (OV)

Table 3. Estimated mean change in day 3 FSH (IU/L) by urinary paraben concentration tertile from linear regression models

Paraben Concentration	N	Estimated Mean Change in FSH (95% CI)[a]	p-value
MP			
Tertile 3 (432-2428 µg/L)	37	0.35 (-1.07, 1.77)	0.63
Tertile 2 (154-430 µg/L)	37	1.04 (-0.39, 2.46)	0.15
Tertile 1 (10.8-144 µg/L)	36	0 (Reference)	
p-value for trend		0.64	
PP			
Tertile 3 (87.8-727 µg/L)	37	1.02 (-0.40, 2.43)	0.16
Tertile 2 (32.0-80.9 µg/L)	37	1.16 (-0.26, 2.57)	0.11
Tertile 1 (0.46-29.4 µg/L)	36	0 (Reference)	
p-value for trend		0.16	
BP			
Tertile 3 (6.00-177 µg/L)	37	0.39 (-1.03, 1.82)	0.59
Tertile 2 (0.96-5.61 µg/L)	37	0.95 (-0.48, 2.39)	0.19
Tertile 1 (<LOD-0.85 µg/L)	36	0 (Reference)	
p-value for trend		0.60	

[a]The parameters estimates can be interpreted as a IU/L change in day 3 FSH for each tertile of paraben urinary concentrations relative to the first tertile (reference). For example, in the 3rd tertile of PP urinary concentrations there is a 1.02 IU/L increase, on average, in day 3 FSH compared to the 1[st] tertile. All model results are adjusted for age. Paraben concentrations are SG-corrected. LOD=0.2µg/L (BP).

Table 4. Estimated mean percent change in AFC by urinary paraben concentration tertile based on Poisson regression model[a]

Paraben Concentration	N	Estimated Mean Percent Change in AFC (95% CI)	p-value
MP			
Tertile 3 (381-2428 μg/L)	47	-10.6 (-28.2, 11.2)	0.31
Tertile 2 (145-377 μg/L)	48	-6.8 (-23.5, 13.7)	0.49
Tertile 1 (5.13-132 μg/L)	47	0 (Reference)	
p-value for trend		0.31	
PP			
Tertile 3 (87.8-727 μg/L)	47	-16.3 (-30.8, 1.3)	0.07
Tertile 2 (26.3-81.8 μg/L)	48	-5.0 (-23.7, 18.4)	0.65
Tertile 1 (<LOD-25.2 μg/L)	47	0 (Reference)	
p-value for trend		0.07	
BP			
Tertile 3 (5.44-177 μg/L)	48	-2.0 (-21.0, 21.6)	0.86
Tertile 2 (0.75-5.12 μg/L)	47	-4.8 (-22.5, 16.8)	0.63
Tertile 1 (<LOD-0.73 μg/L)	47	0 (Reference)	
p-value for trend		0.86	

[a]All model results adjusted for age. Paraben concentrations are SG-corrected. LOD=0.2μg/L (PP, BP).

Table 5. Estimated mean percent change in OV by urinary paraben concentration tertile based on linear regression model[a]

Paraben Concentration	N	Estimated Mean Percent Change in OV (95% CI)	p-value
MP			
Tertile 3 (332-2428 µg/L)	36	4.7 (-19.4, 35.9)	0.73
Tertile 2 (100-326 µg/L)	37	15.0 (-11.2, 49.1)	0.29
Tertile 1 (7.77-95.7 µg/L)	36	0 (Reference)	
p-value for trend		0.73	
PP			
Tertile 3 (87.8-654 µg/L)	36	-0.9 (-23.5, 28.3)	0.94
Tertile 2 (27.1-81.8 µg/L)	37	19.2 (-8.0, 54.4)	0.18
Tertile 1 (<LOD-25.2 µg/L)	36	0 (Reference)	
p-value for trend		0.94	
BP			
Tertile 3 (6.02-102 µg/L)	36	-4.2 (-26.1, 24.2)	0.74
Tertile 2 (0.82-5.43 µg/L)	36	1.0 (-22.2, 31.2)	0.94
Tertile 1 (<LOD-0.80 µg/L)	37	0 (Reference)	
p-value for trend		0.74	

[a]All model results adjusted for age. Paraben concentrations are SG-corrected. LOD=0.2µg/L (PP,